The study of interactions between genome and exposome in the development of systemic lupus erythematosus

Leffers, Henrik Christian Bidstrup; Lange, Theis; Collins, Christopher; Ulff-Møller, Constance Jensina; Jacobsen, Søren

Published in:
Autoimmunity Reviews

DOI:
10.1016/j.autrev.2018.11.005

Publication date:
2019

Document version
Publisher's PDF, also known as Version of record

Autoimmunity Reviews 18 (2019) 382–392

Contents lists available at ScienceDirect

Autoimmunity Reviews

journal homepage: www.elsevier.com/locate/autrev

Review

The study of interactions between genome and exposome in the development of systemic lupus erythematosus☆

Henrik Christian Bidstrup Leffers[a], Theis Lange[b,c], Christopher Collins[d], Constance Jensina Ulff-Møller[a], Søren Jacobsen[a,e,*]

[a] Copenhagen Lupus and Vasculitis Clinic, Center for Rheumatology and Spine Diseases, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark
[b] Department of Public Health, Section of Biostatistics, University of Copenhagen, Denmark
[c] Center for Statistical Science, Peking University, Beijing, China.
[d] Department of Rheumatology, MedStar Washington Hospital Center, Washington, DC, USA
[e] Department of Clinical Medicine, Faculty of Health Science, University of Copenhagen, Denmark.

ARTICLE INFO

Keywords:
Systemic lupus erythematosus
Environment
Exposome
Genetic risk
Interaction
Study design
Study power

ABSTRACT

Systemic lupus erythematosus (SLE) is a systemic inflammatory autoimmune disease characterized by a broad spectrum of clinical and serological manifestations. This may reflect a complex and multifactorial etiology involving several identified genetic and environmental factors, though not explaining the full risk of SLE. Established SLE risk genotypes are either very rare or with modest effect sizes and twin studies indicate that other factors besides genetics must be operative in SLE etiology. The exposome comprises the cumulative environmental influences on an individual and associated biological responses through the lifespan. It has been demonstrated that exposure to silica, smoking and exogenous hormones candidate as environmental risk factors in SLE, while alcohol consumption seems to be protective. Very few studies have investigated potential gene-environment interactions to determine if some of the unexplained SLE risk is attributable hereto. Even less have focused on interactions between specific risk genotypes and environmental exposures relevant to SLE pathogenesis. Cohort and case-control studies may provide data to suggest such biological interactions and various statistical measures of interaction can indicate the magnitude of such. However, such studies do often have very large sample-size requirements and we suggest that the rarity of SLE to some extent can be compensated by increasing the ratio of controls. This review summarizes the current body of knowledge on gene-environment interactions in SLE. We argue for the prioritization of studies that comprise the increasing details available of the genome and exposome relevant to SLE as they have the potential to disclose new aspects of SLE pathogenesis including phenotype heterogeneity.

1. Introduction

Etiologies of autoimmune rheumatic diseases are appreciated to be complex and often multifactorial, involving several genetic factors and exposures to environmental factors in ways that lead to a large phenotypical diversity and remain to be charted. This especially holds true for systemic lupus erythematosus (SLE), which is a clinically heterogenous, multisystem, autoimmune disease characterized by immune dysregulation, serological changes that include a plethora of autoantibodies, immune complex deposition and complement activation, which eventually lead to tissue injury [1].

At least 90% of patients with SLE are women, the incidence in African-American females is > 2.5 times higher than in white females and SLE has the highest incidence rates before the age of menopausal onset [2,3]. The prevalence of SLE varies between 0.02 and 0.1% in European and American populations, with an estimated prevalence of 0.045% among Danes [3–6]. Many genetic and environmental factors have been speculated or shown to influence the varying risks of SLE; most studies have focused separately on each of these factors to increase our understanding of how they may influence the pathogenesis of SLE [7–9]. The existence of familial clustering of SLE is illustrated by the dramatic increase in risk of SLE in close relatives to SLE patients. Relatives of Danish SLE patients were at a 3-fold, 10-fold, 50-fold and 86-fold increased relative risk of SLE in 2nd degree relatives, 1st degree

☆ Supported by grants from the Danish Rheumatism Association (non-commercial) to HCB Leffers (A5094) and S Jacobsen (A3865).

* Corresponding author at: Copenhagen Lupus and Vasculitis Clinic, Center for Rheumatology and Spine Diseases, Rigshospitalet, Section 4242, Blegdamsvej 9, DK-2100 Copenhagen, Denmark.
E-mail address: sj@dadlnet.dk (S. Jacobsen).

https://doi.org/10.1016/j.autrev.2018.11.005
Received 12 November 2018; Accepted 18 November 2018
Available online 14 February 2019

relatives, dizygotic twins or monozygotic twins, respectively [10]. In an Asian population- and registry-based study the degree of relatedness was associated with an even more pronounced relative risk of SLE, where the heritability was estimated to 44% [11]. Genome-wide association studies (GWAS) and observational studies have at present identified > 100 susceptibility genes for SLE, but a meta-analysis of two Asian GWAS could only explain 28% of the heritability [8]. This is well in line with population-based SLE twin studies that show concordance rates of 14%–25% in monozygotic twins [10,12]. This indicates that risk of SLE is influenced by other factors than genetics only, leaving environmental exposures as strong candidates for such other factors.

Environmental factors have by several studies been suggested to significantly impact risk of autoimmune diseases in general [13,14] as well as in SLE [15]. The contribution of environmental factors to the risk of SLE has been estimated to constitute 56% [11]. This estimate does not consider the possibility of biological interactions between the genome and the full spectrum of exposures, the exposome, that may add to the risk of SLE. Further, several exposures have been shown to influence epigenetic mechanism that change gene expression without altering the underlying DNA sequence [9,16]. These observations support the notion that identification of specific gene-environment (G-E) interactions associated with autoimmune diseases including SLE may hold great promise to chart the sources and pathways that underlie these diseases with a view to develop disease prevention and intervention strategies [17].

In SLE, however, such interactions have only been investigated in small cohorts mostly focusing on genes, which are non-specific to SLE, together with a limited number of exposures [18–23].

The aim of this paper is to review the current knowledge on G-E interactions as to the risk of SLE and to address some of the challenges as well as potentials of studying such interactions.

2. Systemic lupus erythematosus

The clinical complexity of SLE is demonstrated by the broad spectrum of clinical manifestations, including cutaneous, musculoskeletal, cardiopulmonary, renal, neurological, and gastrointestinal manifestations. As for laboratory abnormalities, these may reflect hematological and immunological aberrations, which are also included in the syndrome classification of SLE. Immunological aberrations can be reflected by serological changes, including increased levels of autoantibodies and decreased levels of complement in the circulation [24]. At disease onset, the patients may present in clinical clusters that typically are characterized by either acute cutaneous, chronic cutaneous or renal manifestations, i.e. lupus nephritis [25]. During the course of the disease, the incident development of proteinuria as a marker of lupus nephritis may be predicted by lymphopenia and a high number of autoantibodies, including anti-dsDNA antibodies [26,27]. Presence of autoantibodies against phospholipids and/or the lupus anticoagulant delineate another subset of patients with increased risk of thrombotic events [1,28].

Identifying the time of disease onset remains a crucial challenge as this may be of pivotal significance in the discrimination between disease triggering factors and factors that may modify the course of the disease. It is well recognized that in persons who subsequently develop SLE, serological changes may predate clinical manifestation of the disease by > 9 years [29,30]. The early autoantibody repertoire does not epitope-wise seem to differ substantially from the repertoire of established disease, i.e. comprising mainly autoantibodies against cellular and nuclear components, but the number of autoantibodies does increase during the preclinical phase [29] and the activation of the type I interferon (IFN) system has been shown to predict manifest autoimmune connective tissue disease in subjects with anti-nuclear antibodies [31].

Activation of the type I IFN system in SLE patients can be initiated by signaling through Toll-like receptors (TLR) and other nucleic acid-sensing mechanisms, leading to the production of immune complexes, looped activation of persistent type I IFN responses, and neutrophil activation and trapping [1,32]. Endogenous sources of TLR triggers have been linked to impaired clearing of cellular and nuclear debris derived from various types of cell death, including apoptosis, due to various deficiencies of innate immunity [33]. However, it is of interest that not only extracellular vesicles derived from apoptotic nucleated cells may have pro-inflammatory effects, but also vesicles shed from activated platelets [34] with the latter having a rich content of mitochondria containing DNA similar to bacterial DNA due to its high content of CpG-regions [35]. It does, however, remain to be fully determined, to which extent environmental agents may act as direct triggers of immune activation and/or have modulating effects hereon in SLE patients.

Aberrations of innate immunity as well as adaptive immunity may also impact the clinical course of disease with respect to e.g. nephritis, thrombosis and infections [36]. Understanding the early events of SLE may not only provide pathogenetic insight but may also offer opportunities to attenuate the development of comorbidities, organ damage and increased mortality attributable to SLE [37] by improving diagnostics and avoiding therapeutic delay [38].

2.1. Genetic factors modifying risk of systemic lupus erythematosus

The promising outlook of studying the genome for genetic contribution and risk variants for SLE [8] has been challenged by the fact that the multitude of identified susceptibility genes confer such low risk that they do not provide sufficient discrimination to allow rational use in clinical practice. As for the more strongly associated variants, representing for example deficiencies of early complement components, these occur with such rarity that they are seldom encountered in clinical practice. Combinations of genes present in the same individual may through gene-gene epistatic interaction confer increased risk of SLE beyond the risk associated with the individual genes, which also holds true for some gene-sex interactions [39]. The clinical implications of such genetic findings, including the genetic distinction between various SLE phenotypes are still in wanting [40]. However, detection of new variants by whole-genome sequencing and imputations may allow a substantial increase in the number of recognized common and rare sequence variants, which as a whole may associate with specific phenotype traits as demonstrated for human serum immunoglobulin levels [41].

Mendelian or monogenic SLE comprises a few rare variants in genes such as *C1Q* (complement 1q deficiency) or *TREX1* (downregulation of exonuclease that degrades single- and double-stranded DNA) [42,43]. As suggested by Terruel et al. [44], other rare variants and more commonly occurring gene variants associated with SLE can be grouped into clusters that relate to pivotal elements of SLE pathogenesis. Such clusters can range functionally from innate to adaptive immunity and homeostatic cellular functions (Fig. 1A).

Autoantigen triggering of TLR 7 by ssRNA nucleotides activates type I IFN signaling and homozygous variation of *TLR7* (rs3853839) is associated with SLE (2.9% in controls vs. 7.6% in SLE patients in European populations) and even more with lupus nephritis (14%) [45,46].

Type I IFN signaling in myeloid and T cells relies in part on signal transducer and activation of transcription 4 (STAT4) and is also central in the differentiation of Th1 and Th17 responses [47]. Variation of *STAT4* (rs7574865, frequency of 22% in controls vs. 31% in SLE patients in the north American population) [47] has been associated with SLE as well as cardiovascular morbidity in such patients [48,49].

Protein tyrosine phosphatase non-receptor 22 (PTPN22) plays a key role in several aspects of T cell function and variations of *PTPN22* (rs2476601, frequency of 8% in controls vs. 11% in SLE patients in the European population) has been associated with SLE [50].

The major histocompatibility complex (MHC) belongs to the most

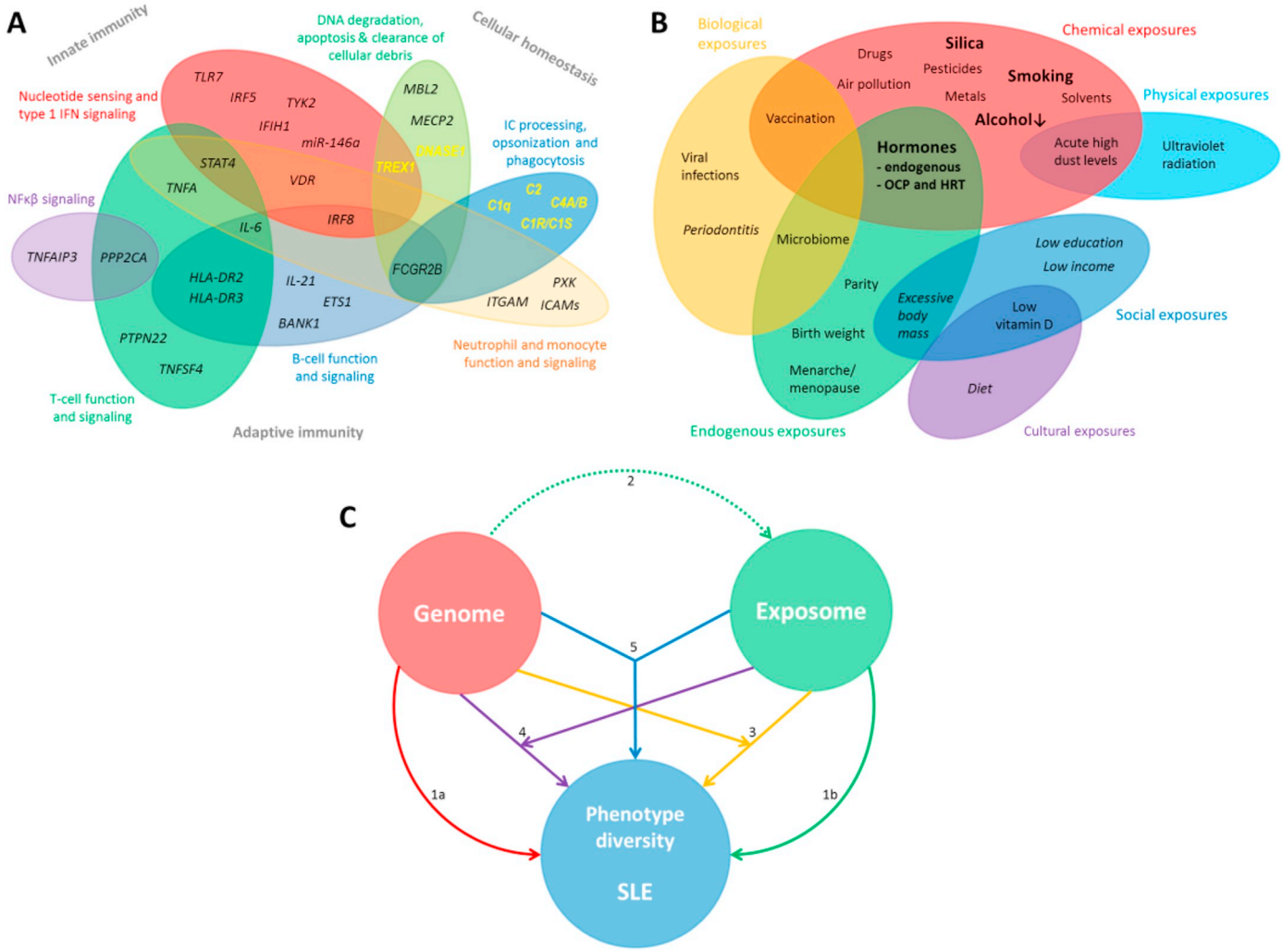

Fig. 1. Genetic and environmental factors influencing the risk of systemic lupus erythematosus (SLE) and how such factors may interact.
A: Genes associated with risk of SLE clustered by which immunobiological functions they may affect, modified from Teruel et al. [8], surviving two different Bayesian approaches (false positive report probability and Bayesian false discovery probability from a meta-analysis of observational studies by Jeong et al. [112]). Genes in yellow display rare variants with high prevalence of SLE.
B: Exposures with known or suggested association with SLE. Bold letters: Exposures with confirmed association with SLE. Plain letters: Exposures suggested to increase risk of SLE but not confirmed [15]. In italics: Other exposures of interest, which have been demonstrated to influence risk of other systemic autoimmune diseases, i.e. rheumatoid arthritis [93,113].
OCP: Oral contraceptive pill, HRT: Hormone replacement therapy.
C: Framework describing plausible models of interaction between genetic factors and environmental exposures modifying disease risk [85]. Red (1a) and green (1b) arrows: Genotype and environmental factor influence disease risk independently. Dotted green arrow (2): Genotype changes exposure of environmental factor. Yellow arrows (3): Genotype changes the expression or impact of the environmental exposure. Violet arrows (4): Environmental factor changes the risk effect of genotype. Blue arrow (5): Genotype and environmental factor are both required to influence disease risk or phenotype. Numbers in parenthesis correspond with the numbering in Section 3.1.

polymorphic regions of the genome, and is strongly associated with SLE [51], as also observed in other autoantibody mediated diseases such as rheumatoid arthritis (RA). In European [52] and North American [53] populations extended haplotypes susceptible to SLE are HLA-DR2 and HLA-DR3. Twelve percent of cases vs. 21% of SLE patients and 8% of controls vs. 17% of cases in a North American population carry these haplotypes, respectively [53].

2.2. Exposures modifying risk of systemic lupus erythematosus

The concept of integrating life-course exposures to an individual, from the prenatal period and onwards, and terming such exposures as an exposome was first published in 2005 [54], and has later been suggested to encompass the result of the combined exposures from all sources that reach the internal chemical environment of an individual [55]. Combined with traditional epidemiological classification of potential external hazards, such as physical, chemical, biological, social and cultural factors, the concept of an exposome provides a more complete framework for the understanding and conceptualization of exposures, being exogenous or endogenous. Such a framework might improve the overview and understanding of potential roles of various exposures relevant to the development of SLE as tentatively depicted (Fig. 1B).

Several environmental exposures have been suggested to be associated with the development of SLE. Current smoking, exposure to crystalline silica and intake of exogenous hormones comprise the strongest increase in relative risk by a factor 1.5, 2.1–4.6 and 1.5–1.9, respectively [56–58]. Light to moderate alcohol consumption seems to be inversely associated with SLE [59]. To this end it is of interest that smoking may exert pro-inflammatory effects on some parts of

immunopathogenesis and anti-inflammatory effects on other parts hereof [60].

Pesticides, chemical and industrial exposures have been suggested to increase risk of SLE by a factor of 1.5–7.4 in various settings [61]. Metabolomic studies have demonstrated a close association between welding fume exposure and systemic inflammation [62]. It has also been speculated that several infectious agents may trigger the development of SLE and increase the severity of the clinical phenotype; although EBV-associated infectious mononucleosis was not associated with SLE in a registry-based population study [63,64]. Other evidence suggest a lack of control of Epstein-Barr virus (EBV) infection in SLE patients [65]. In clinical practice, it is well-known that ultraviolet (UV) radiation may exacerbate existing SLE disease and provoke characteristic skin changes in photosensitive patients [66]. However, the etiological or pathogenic roles of the latter remain to be resolved [66] and to this part, several other components of the exposome should be considered and explored as potential factors modifying the risk of SLE.

2.3. Bridging the gap between genome and exposome

The cross-field between the genome and the exposome is enormous but of even greater interest as it has still not been possible to single out any environmental or genetic factors that come just near defining a comprehensive risk model of SLE. However, environmental factors may interact with genetics by several mechanisms, which may have the potential to add significantly to such wanted risk models. Epigenetics are stable and heritable, yet reversible, mechanisms that regulate gene expression without altering the underlying gene code [67]. One key epigenetic mechanism is DNA methylation, which controls the accessibility to gene regulatory regions. Aberrant DNA methylation patterns have been increasingly recognized in SLE patients compared to healthy controls in epigenome-wide association studies (EWAS). These have consistently suggested a pattern of DNA hypomethylation of IFN-regulated genes, such as *IFI44L, PARP9, IFITM1* and *MX1*, in $CD4^+$ T cells, monocytes, granulocytes and B cells [68–72], as well as in white blood cells (WBCs) in general [73,74]. Differential methylation of WBCs in SLE patients has even been suggested to be associated with the presence of SLE-related autoantibodies [75]. Recently, we have further shown that B cells from Danish SLE twins additionally exhibit a pattern of promoter hypermethylation best explained by *TNF* and *EP300* as upstream regulators, the significance of which is still unknown [72] DNA methylation is known to vary by sex [76] and ethnicity [77] and may further be influenced by smoking [78] and nutrition [79]. Interestingly, the DNA methylation patterns of monozygotic twins have been shown to diverge throughout life [80], providing further evidence that epigenetic factors may contribute to the phenotypic variation observed in monozygotic twins [72]. Transcription of non-coding RNAs (microRNAs) from intronic regions, which regulate gene expression and posttranslational modification of histones e.g. acetylation are other examples of epigenetic factors that are aberrantly expressed in SLE [81,82].

Recently, it has been demonstrated that EBV produced EBNA2 proteins occupy about half of the SLE susceptibility loci and co-clusters with transcriptional factors, thereby altering gene expression [83]. This very intriguing finding implies a new potentially pathogenic role of EBV in SLE, and the finding need to be corroborated by controlled transcriptomic studies.

2.4. Rationale for studying gene-environment interaction in SLE

Current knowledge indicates a multifactorial risk profile of SLE including a multitude of gene variants and several environmental exposures that, however, do not fully account for the total risk of SLE. The identification of gene-gene interactions and development of gene risk scores have increased our predictive capacity but still not to an extent that has direct clinical implication. Unravelling key environmental triggers in distinct genetically susceptible individuals may provide new insight into SLE pathophysiology needed for development of new preventive and therapeutic strategies. Although it seems intuitively obvious that such G-E interactions may contribute to the total risk of SLE, our current understanding of such interactions is limited as well as the available data needed for us to reach such an understanding. There is therefore a large unmet need for studies that comprise genomic and phenotypic data as well as comprehensive exposure data reaching far back in time before onset of any subclinical immunological aberrancies [13]. Encountering this need calls for complex and large datasets specifically designed to address G-E interactions.

3. Definition and measures of gene-environment interaction

3.1. Gene-environment interaction

In general, studies of G-E interactions aim to describe how genetic and environmental factors jointly influence the risk of developing human disease [17]. This includes differential effects on disease risk of an environmental exposure in persons with different genotypes and vice versa [84]. To conceptualize relationships between genetic and environmental factors and their effect on disease risks, Ottman [85] described five plausible, not necessarily exhaustive, models of such relationships (Fig. 1C); 1a) genotype and 1b) environmental exposures influence disease risk independently of each other, e.g. variants of *STAT4* and smoking are both established risk factors for SLE, but have not yet been shown to display interaction in this respect [21]; 2) the genotype changes the amount of exposure of an environmental factor, e.g. common variant in the nicotine acetylcholine receptor gene cluster associated with nicotine dependency associates with lung cancer [86]; 3) the genotype changes the impact of a given environmental exposure, e.g. higher risk of SLE in smokers who are non-rapid acetylators [20]; 4) the exposure changes the effect of the genotype, e.g. the influence of diet on expression of autoimmune-associated genes and disease severity by epigenetic mechanisms in a murine model of SLE [87]; 5) genotype as well as exposure are required to raise risk, e.g. development of hemolytic anemia in glucose-6-phosphate dehydrogenase deficiency genotype and exposure to certain medications that not otherwise would cause hemolytic anemia [88]. Each of these models may serve as scaffolds in the unravelling of SLE pathogenesis by classifying potential interactions between genetic factors and exposures.

3.2. Interaction

Interaction between two factors is present if the simultaneous occurrence or effect of these factors changes the risk of disease from what would be expected by the effect of each of the factors themselves. The above described models thus describe relationships that are respectively 1) non-interacting by definition, 2–4) interacting and 5) dependent on occurrence of both factors. Epidemiological studies combining genetic data with exposure data may provide clues for evaluating the probability of biological interaction by estimating interaction statistically. Based on disease rates, i.e. from cohort studies, it is demonstrated that independent risk factors adhere to an additive risk model and that biological interaction results in departure from additivity of disease rates by risk factors [89].

Case-control studies without the dimension of time can instead provide multiplicative odds ratios. Departure from additivity or multiplicativity, may both indicate biological interaction and various measures of such interaction may be calculated depending on the statistical model employed [90].

To determine biological interaction in widely used statistical models, forehand assumption of additive or multiplicative relationship between independent risk factors is needed and it should be possible to describe sufficient causes as the complete set of causal mechanisms that may contribute to causing disease [91]. Sufficient cause is most often

Table 1
Statistical methods and measures to indicate biologic interaction.

	Measure	Interpretation	Null hypothesis
Additive interaction	RERI: relative excess risk due to interaction	Additional risk compared to the expected risk from adding the risk for each exposure	RERI = 0
	AP: Attributable proportion due to interaction	Proportion due to interaction of overall risk among those with both exposures	AP = 0
	S: Synergy index	Excess risk from combined exposures relative to the risks from each exposure	S = 1
Multiplicative interaction	ROR: Ratio of odds ratios	Additional risk compared to the expected risk from multiplying risks of each exposure	ROR = 1
Other methods of indicating interaction	Comparison of likelihood ratio statistics between models with and without interaction term	The model with the interaction term is a better fit with the data, and thus that interaction should be considered	Models are equally good
	Generalized multifactor dimensionality reduction. Test of significance of identified model.	A measure of the degree of consistency with which the selected interaction is identified as the best model among all possibilities considered	Models are equally good

not a single factor, but a minimum set of factors and circumstances that, if present in a given individual, will cause the disease [89].

3.3. Measures of biological interaction

Most previous studies of SLE risk association have used a case-control design, in which the number of cases and controls are determined by investigators and time is not a part of the equation. Here, odds ratios can be calculated by a variety of methods including logistic regression analysis. Odds ratios can under the rare disease assumption approximate relative risks [92], thus allowing approximation of additive interaction measures (Table 1), including relative excess risk (RERI) and attributable proportion (AP) due to interaction as well as the interaction synergy index (S) as previously described [93]. Multiplicative interaction can be expressed as the ratio of odds ratios (ROR) between the observed odds ratio for combined exposure compared to that expected by multiplying odds ratios of each exposure. Other ways to determine statistical interaction include the comparison of statistical models with and without the interaction in question testing the null hypothesis that the models perform equally well.

4. Studies of gene-environment risk interactions in systemic lupus erythematosus

Given the rationale for identifying potential G-E interactions in SLE, surprisingly few studies have addressed this. Some G-E interactions in SLE have been suggested, although these have mainly been suggested by studies of G-E interactions in other diseases [94]. These comprise interactions between certain environmental factors and variations in genes encoding xenobiotic-metabolizing enzymes, e.g. *N*-acetyltransferase 2 (*NAT2*) and glutathione S-transferase M1 (*GSTM1*), as well as the protein coding gene of estrogen receptor alpha (*ESR1*). Also, two genes with more direct immunological significance have been investigated, namely the candidate risk gene, *STAT4*, alongside *TNFRSF1B* as summarized below and in Table 2.

4.1. Slow acetylation activity - smoking and dietary habits

Studies in the 1970s revealed a predominance of individuals with slow acetylation activity among patients with drug-induced SLE [95,96]. In a case-control study of 152 Japanese SLE-patients, the modifying effect of polymorphisms in the *NAT2* gene on history of smoking, alcohol- and caffeine-rich beverage intake was investigated [21,23]. Subjects were divided by genotype into rapid acetylators (*NAT2*4* homozygous) and non-rapid acetylators (*NAT2*4* non-homozygous). One analysis suggested an increase in risk of SLE in smoking subjects that were non-rapid acetylators by an odds ratio of 6.44 compared to the non-smoking rapid acetylators with an AP of 0.5, indicating that half of the total added risk studied was explained by interaction. However, the ROR was not statistically significant. This discrepancy might be explained by the faulty assumption that odds ratios may approximate the relative risk in that study. In further analyses, a history of alcohol consumption was found to interact with non-rapid acetylation associating negatively to SLE, whereas non-rapid acetylators who had a history of drinking black tea were positively associated with SLE. No interactions were detected for drinking coffee or black tea.

4.2. Xenobiotic metabolization - smoking and sun exposure

In the above mentioned Japanese population, variants in other genes encoding enzymes responsible for metabolizing xenobiotics were also investigated, such as *CYP1A1* encoding a cytochrome P450 monooxygenase and *GSTM1*. Phase 1 enzymes, such as CYP1A1 metabolically activate carcinogenic or toxic substances in cigarette smoke to highly reactive electrophiles that can be detoxified by phase 2 enzymes such as glutathione S-transferases. Balance between phase-1 and -2 enzymes could therefore be speculated to influence select elements of the exposome relevant to the development of SLE. In that study, the AP was 0.60 for ever-smoking combined with *CYP1A1* rs4646903. However, interaction on a multiplicative scale was not observed. No interaction measures between history of smoking and *GSTM1* null were found to be statistically significant. Combining these two risk genotypes with smoking provided an AP of 0.68 to the risk of SLE, but again no interaction on a multiplicative scale was observed [19].

Another case-control study, comprising 243 SLE cases from North- and South Carolina, USA, investigated three enzymes of the glutathione S-transferase family (*GSTM1, GSTT1* and *GSTP1)* as to their possible interaction with sun exposure [22]. Sun exposure was determined as occupational sunlight exposure defined as working > 24 months in a sun exposed job. Overall, no interactions were detected. However, in a subset of Caucasians with *GSTM1* null, an OR of 3.1 (95% CI: 0.9–10.8) was determined. A likelihood ratio test of statistical models with and without the interaction was not able to retain the null-hypothesis, for which reason it was concluded that there might be an interaction. Although the effect size of this interaction was not stated, it must be expected to be low based on the data.

4.3. Estrogen receptor alpha - smoking and alcohol

Given that 80–90% of SLE occurs in women, variants of the gene encoding estrogen receptor alpha, *ESR1*, have also been investigated in SLE [18]. Variations in *ESR1* have been studied in many diseases and found to be associated with e.g. breast cancer and endometrial cancer [97]. In a Chinese case-control study comprising 230 SLE patients, the combined presence of *ESR1* rs2234693 C-allele and current smoking history was positively associated with SLE with an odds ratio of 2.5 (Table 2). Presence of interaction was determined by comparing models

Table 2
Published studies of gene-environment (GxE) interactions in the development of systemic lupus erythematosus (SLE).

Publication	Population/sample	Genetic analysis and assessment of environmental exposure	Rationale behind proposed GE interaction	Specific interactions - G × E investigated	Odds Ratio - G × E	Results by available measures of interaction (95%CI)					
						Additive interaction measures			Multiplicative interaction measure ROR	Non-parametric tests	
						RERI	AP	S		Likelihood ratio test[a]	GMDR analysis
Fraser et al. 2002 [22]	US. The Carolina Lupus study. Cases: 243 Controls: 298	PCR. Structured interview (60 min.) regarding outdoor work-occupancy.	The GST genes catalyze metabolic pathways for the excretion of reactive oxygen species induced by UV-radiation in sunlight	*GSTM1* null × sunlight (Caucasians: 39 cases, 92 controls)	1.6 (0.5–1.5)	–	–	–	–	**p = .028**	**–**
				GSTT1 null × sunlight	Not reported	–	–	–	–	Interaction not found	–
				GSTP1 Val/Val or Val/Ile × sunlight	Not reported	–	–	–	–	Interaction not found	–
Kiyohara et al. 2009 [20]	Japan. Kyushu University Hospital - Saga University hospital. Sapporo Medical University Hospital. Cases: 152 Controls: 251 or 457	PCR-RFLP. Self-administered questionnaire.	NAT2 detoxifies aromatic amines, an important carcinogen from smoke. Variant alleles cause slower rate of detoxification.	*NAT2* non-*4* homozygous (non-rapid acetylator genotype) × history of smoking	6.4 (3.1–13.5)	3.3 (−0.5–7.1)	**0.5 (0.1–0.9)**	2.4 (0.8–7.0)	1.44 (0.5–4.6)	–	–
Kiyohara et al. 2009 [21]			STAT4 is considered as part of IFNα pathway and a key molecule in Th1 and Th17 lineages playing an important role in chronic inflammatory disorders and has been shown activated by smoking.	*STAT4* rs7574865 × history of smoking	4.4 (2.2–8.6)	1.4 (−1.5–4.2)	0.3 (−0.2–0.8)	1.7 (0.6–4.6)	1.2 (0.5–3.3)	–	–
			TNF plays a crucial role in a wide variety of proliferative responses, inflammatory effects and immune responses. TNFRSF1B has also been proposed to play a role in apoptotic mechanisms.	*TNFRSF1B* rs1061622 × history of smoking	5.4 (2.5–11.8)	2.65 (−1.35–6.65)	**0.49 (0.07–0.92)**	2.54 (0.8–8.1)	1.7 (0.6–4.6)	–	–
Kiyohara et al. 2012 [19]			Phase 1 enzymes, such as CYP1A1 metabolically activate carcinogenic or toxic substances in cigarette smoke to highly reactive electrophiles that, phase 2 enzymes such as GSTs, can be detoxified. Balance between phase 1 and − 2 enzymes may be an important.	*CYP1A1* rs4646903 × history of smoking	9.7 (2.7–34.6)	5.9 (−6.3–18.0)	**0.6 (0.1–1.1)**	3.0 (0.7–13.9)	1.7 (0.4–7.4)	–	–
				GSTM1 null × ever smoking	3.4 (1.8–3.7)	1.0 (−1.2–3.3)	0.3 (−0.2–0.9)	1.8 (0.5–6.7)	1.5 (0.58–3.7)	–	–
				CYP1A1 homozygosity combined with *GSTM1* null × ever smoking	17.5 (3.2–96.0)	12 (−17.7–41.8)	0.7 (0.11–1.25)	3.6 (0.5–25.4)	1.7 (0.2–11.2)	–	–

(*continued on next page*)

Table 2 (*continued*)

Publication	Population/sample	Genetic analysis and assessment of environmental exposure	Rationale behind proposed GE interaction	Specific interactions - G × E investigated	Odds Ratio - G × E	Results by available measures of interaction (95%CI)					
						Additive interaction measures			Multiplicative interaction measure ROR	Non-parametric tests	
						RERI	AP	S		Likelihood ratio test[a]	GMDR analysis
Kiyohara et al. 2014 [23]			NAT2 is an important xenobiotic-metabolizing enzyme, impaired ability to remove reactive substances e.g. caffeine from the body may play a role in the etiology of SLE.	*NAT2*4* non-homozygous (non-rapid acetylator genotype) × ever alcohol	0.2 (0.1–0.7)	–	–	–	–	**p = .026**	**–**
				*NAT2*4* non-homozygous (non-rapid acetylator genotype) × ever green tea	0.8 (0.4–1.9)	–	–	–	–	p = .16	–
				*NAT2*4* non-homozygous (non-rapid acetylator genotype) × ever black tea	3.6 (1.5–8.2)	–	–	–	–	**p = .048**	**–**
				*NAT2*4* non-homozygous (non-rapid acetylator genotype) × ever coffee	2.1 (0.9–5.0)	–	–	–	–	p = .094	–
Zhou et al. 2017 [18]	China. Affiliated Hospital of Qingdao University. Cases: 230 Controls: 462	PCR-RFLP. Questionnaire administered by trained staff.	ESR1 interacts with transcription factors, coactivators and co-repressors regulating gene expression. ESR1 gene polymorphisms have been described as associated with various complex diseases, however no interaction with alcohol nor smoking has been determined.	*ESR* rs2234693 × smoking	2.5 (1.7–3.4)	–	–	–	–	–	**p = .001**
				ESR1 rs2234693 × alcohol	Not reported	–	–	–	–	–	p = .06

P-values numbers in bold indicate a significant deviation from null-hypothesis ($p < 0.05$)

GST: glutathione S-transferase, NAT2: *N*-acetyl-transferase 2, STAT4: signal transducer and activator of transcription 4, TNFR: tumor necrosis factor receptor, CYP: cytochrome P450, PCR: Polymerase chain reaction, RFLP: restriction fragment length polymorphism, GMDR: generalized multifactor dimensionality reduction, RERI: relative excess risk due to interaction, AP: attributable proportion due to interaction, S: synergy index, ROR: ratio of odds ratios.

[a] Comparison of likelihood ratio statistics with and without the interaction in question.

in a generalized multifactor dimensionality reduction. As for the likelihood ratio test, the effect size of the interaction could not be quantified. Only a statistically non-significant tendency towards interaction between this allele and alcohol was observed [18].

4.4. STAT4 and TNFRSF1B variants - smoking

Variations in *STAT4* and *TNFRSF1B* have been studied in the same Japanese cohort of 152 SLE patients mentioned above as to their potential interaction with smoking. TNF plays a crucial role in a wide variety of proliferative responses, inflammatory effects and immune responses and tumor necrosis factor receptor superfamily, member 1B TNFRSF1B has furthermore been proposed to play a role in apoptotic mechanisms [98]. The central role of STAT4 as part of type 1 IFN signaling is well established and these genes were therefore of interest to study, however, it was not stated which specific mechanism of interaction that was hypothesized. The AP was found to be 0.49 for *TNFRSF1B* rs1061622 in combination with history of smoking, but presence of interaction was not supported in a multiplicative model. Nor did *STAT4* rs7574865 seem to interact with history of smoking [21].

5. Challenges in determining gene-environment interactions in SLE

5.1. Causation

Causation of disease has multiple definitions in the epidemiologic literature and are suggested to comprise production of effect, necessary cause, sufficient cause, probabilistic cause and counterfactual cause [99]. Cause of disease has widely been defined as an event, condition or characteristic that preceded the disease event and without which the disease event either would not have occurred at all or would not have occurred until some later time [100]. For epidemiological associations to be indicative of causation, they should be supported by observations that align with principles of causation as for example Hills criteria for infection [101] or by direct experimental study of disease mechanisms. Furthermore, a given disease can be caused by more than one causal mechanism, which may involve the joint action of a multitude of components; the complexity of which, renders the study of most biological effects unresolved [91]. As mentioned earlier, tobacco smoking is associated with SLE, but is by itself not a sufficient cause as smoking will not cause SLE in everyone [100]. Given the currently obscure and multifactorial nature of the pathogenesis of SLE, it does not seem feasible to identify the complete set of causal mechanisms that sufficiently explain the disease risk [91]. In complex disease it therefore seems more feasible to operate with a probabilistic cause definition [99]. Making causal inferences based on epidemiological studies of G-E interactions limited to only conveying associations may seem futile. However, the experience in RA, where epidemiological observations of increased risk conferred by smoking exposure and presence of distinct MHC haplotypes lead to the unravelling of a central disease mechanism of seropositive RA serves as an encouraging story. Peptide citrullination induced by smoking, restriction by HLA-DR shared epitope and production of autoantibodies against citrullinated peptides (ACPA) are now well-established components of RA pathogenesis [102,103]. Assuming causality, the AP of smoking among carriers of HLA-DR shared epitope to the risk of ACPA positive RA is estimated at 36% [104].

5.2. Study design

Cohort studies, in which large populations are followed prospectively, may provide concise information on common exposures and demographic factors before individuals develop overt clinical disease. This offers low selection and recall bias and provides direct estimates of incidence rates. However, for rare sequence variants and rare phenotypical presentations as is the case in SLE, very large cohorts are required to achieve adequate statistical power. Case-control studies may thus provide a more direct approach towards studying G-E influences and interactions on the risk of SLE. The case-control design offers the opportunity to model and size case-control groups and even to include relatives of cases as controls in the design [85]. In spite of pitfalls such as susceptibility to recall bias and selection bias [93] the possibility to impute even long haplotypes for close and distant non-genotype relatives in large sample sets may convert the case-control data set to allow cohort type of analyses [105].

5.3. Study power

Sample-size requirements for G-E studies can be enormous as detection of a statistical interaction may require four times the sample size for detecting a main effect of comparable magnitude [94,106]. Sample sizes of thousands of cases are typically needed for interaction analyses in candidate-gene studies, and tens of thousands may be needed in GWAS, because of the stringent significance levels required to overcome inflated type 1 error rates using various correction methods, such as Bonferroni correction or false discovery rates [106,107]. It might be speculated whether these conservative correction approaches combined with the rapidly increasing amount of data produced by various omics technologies and data sources may prevent us from making new discoveries. This highlights the need for new analytic approaches to “big data” while considering the opportunity to increase the precision by which phenotype and disease classification can be made; hereby power of G-E interaction studies may be increased by better case discrimination [108].

5.3.1. Increasing statistical study power

Key determinants of statistical power in case-control studies of G-E interaction are allele frequency, exposure frequency, odds ratios for the main effects, the magnitude of interaction to be detected, choice of acceptable levels for type 1 and 2 errors, and sample size [109]. In diseases where case sample size may be limited due to rarity of the disease as is the case in SLE, statistical power of such studies can be increased by increasing the ratio of controls against cases [110]. To illustrate this, we performed simulations of statistical power [111] of a fictive case-control study aiming to detect statistical interaction between a *STAT4* variant and exposure to smoking on the risk of SLE,

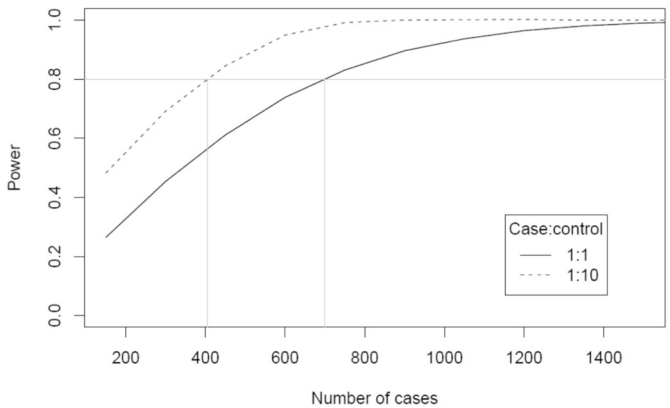

Fig. 2. Statistical power in simulated case-control studies of a gene-environment interaction. Power simulations were performed using the following settings: estimated odds ratio for SLE of 1.5 for current smoking (prevalence of 22%, The Danish Health Authority, 2018); estimated odds ratio for SLE of 1.5 for carriage of polymorphism of *STAT4* (prevalence of 31% [47]); ratio of odds ratios = 2; type 1 error = 0.05. Unbroken and dotted lines represent simulation products for case:control ratios of 1:1 and 1:10, respectively. The vertical lines indicate the corresponding number of cases needed to achieve 80% power (horizontal line). MAF: Minor allele frequency.

(Fig. 2). The upper and lower curves representing 1:1 and 1:10 case:-control ratios, respectively, clearly demonstrate how the increase in number of controls left-shifts the curves towards higher statistical power. In this simulation, increasing the number of controls by a factor 10 more than halved the number of cases needed; alternatively, this approach may also allow regulation of other determinants of statistical power towards more critical levels.

6. Conclusions and future directions

Over the past several years there has been increasing progress made in elucidating potential genetic and environmental factors, which contribute to the development of SLE in individual populations. GWAS studies have identified numerous candidate susceptibility genes and large cohort studies have suggested several environmental exposures, which may influence disease risk. However, no single gene (or combination of genes) nor the spectrum of exposures of an individual has yet emerged, which can account for more than a minor fraction of those who go on to develop SLE. The limited ability of common variants to account for the genetic contribution to complex disease has prompted searches for rare gene variants with large cumulative effects, to partly assess components of the currently obscure part of genetic heritability.

Scientists have long postulated that it is a combination of both genetic and environmental risk factors that confer the greatest risk for SLE development, and potentially the most influential mechanisms contributing to SLE risk is through synergistic G-E interactions. The study of G-E interactions in SLE also holds an enormous potential to dissect into the pathophysiologic processes of SLE and may pinpoint areas of interest for further exploration of epigenetics and gene expression studies as for example suggested for IFN signaling in SLE. Given what we know about G-E interactions in other autoimmune conditions, it is credible to assume such interactions exist that influence SLE risk and pathogenesis. The known profound risk association for RA in smokers with distinct MHC haplotypes has provided significant understanding into disease pathogenesis as well as providing a potential for RA prevention in high risk individuals. As a promising outlook, G-E interaction discoveries may have the potential to provide valuable knowledge into modifiable risk behaviors, which contribute to disease risk in predisposed individuals.

The broad heterogeneity of SLE, however, remains a significant challenge to discovering such G-E interactions as does the relative rarity of the disease in the general population. Studies that have attempted to look at candidate G-E interactions in SLE have mostly all suffered from lack of statistical power due to the relative rarity of specific genetic factors in combination with defined environmental exposures. Determining the existence of any such interactions thus calls for alternate and optimized models of discovery. As outlined in this paper, one such approach is through case-control studies where statistical power is increased by increasing the ratio of controls to cases. Such strategies, as well as ongoing efforts to better subclassify specific clinical phenotypes of SLE and determine more homogeneous case populations, may provide more clear insights to specific G-E interactions, which influence disease development. As continued emerging candidate genes and combinations hereof become better elucidated along with specific exposal patterns, then defined SLE subset populations could be examined in novel designed case-control studies such as here proposed.

Faced with the complexity of SLE, the need to combine epidemiological and genetic research methods is apparent and can help to provide better insight into complex pathway mechanisms accounting for disease heterogeneity and different phenotypes. Such insight might even have the potential to form base of new approaches for disease classifications based on specific molecular pathology rather than crude clinical presentations. Pointing out key pathophysiological processes would also indicate future directions for the development of new diagnostic tools and personalized therapeutic approaches.

Acknowledgements

None.

References

[1] Kaul A, Gordon C, Crow MK, Touma Z, Urowitz MB, van Vollenhoven R, et al. Systemic lupus erythematosus. Nat Rev Dis Primers 2016;2:16039.
[2] McCarty DJ, Manzi S, Medsger Jr. TA, Ramsey-Goldman R, LaPorte RE, Kwoh CK. Incidence of systemic lupus erythematosus. Race and gender differences. Arthritis Rheum 1995;38(9):1260–70.
[3] Hermansen ML, Lindhardsen J, Torp-Pedersen C, Faurschou M, Jacobsen S. Incidence of systemic lupus erythematosus and lupus nephritis in Denmark: a Nationwide cohort study. J Rheumatol 2016;43(7):1335–9.
[4] Furst DE, Clarke AE, Fernandes AW, Bancroft T, Greth W, Iorga SR. Incidence and prevalence of adult systemic lupus erythematosus in a large US managed-care population. Lupus 2013;22(1):99–105.
[5] Nived O, Sturfelt G. Epidemiology of systemic lupus erythematosus. Monogr Allergy 1987;21:197–214.
[6] Rees F, Doherty M, Grainge MJ, Lanyon P, Zhang W. The worldwide incidence and prevalence of systemic lupus erythematosus: a systematic review of epidemiological studies. Rheumatology (Oxford) 2017;56(11):1945–61.
[7] Barbhaiya M, Costenbader KH. Environmental exposures and the development of systemic lupus erythematosus. Curr Opin Rheumatol 2016;28(5):497–505.
[8] Teruel M, Alarcon-Riquelme ME. The genetic basis of systemic lupus erythematosus: what are the risk factors and what have we learned. J Autoimmun 2016;74:161–75.
[9] Richardson B. The interaction between environmental triggers and epigenetics in autoimmunity. Clin Immunol 2018;192:1–5.
[10] Ulff-Moller CJ, Svendsen AJ, Viemose LN, Jacobsen S. Concordance of autoimmune disease in a nationwide Danish systemic lupus erythematosus twin cohort. Semin Arthritis Rheum 2018;47(4):538–44.
[11] Kuo CF, Grainge MJ, Valdes AM, See LC, Luo SF, Yu KH, et al. Familial aggregation of systemic lupus erythematosus and Coaggregation of autoimmune diseases in affected families. JAMA Intern Med 2015;175(9):1518–26.
[12] Deapen D, Escalante A, Weinrib L, Horwitz D, Bachman B, Roy-Burman P, et al. A revised estimate of twin concordance in systemic lupus erythematosus. Arthritis Rheum 1992;35(3):311–8.
[13] Bogdanos DP, Smyk DS, Invernizzi P, Rigopoulou EI, Blank M, Sakkas L, et al. Tracing environmental markers of autoimmunity: introducing the infectome. Immunol Res 2013;56(2–3):220–40.
[14] Floreani A, Leung PS, Gershwin ME. Environmental basis of autoimmunity. Clin Rev Allergy Immunol 2016;50(3):287–300.
[15] Parks CG, de Souza Espindola Santos A, Barbhaiya M, Costenbader KH. Understanding the role of environmental factors in the development of systemic lupus erythematosus. Best Pract Res Clin Rheumatol 2017;31(3):306–20.
[16] Canas CA, Canas F, Bonilla-Abadia F, Ospina FE, Tobon GJ. Epigenetics changes associated to environmental triggers in autoimmunity. Autoimmunity 2016;49(1):1–11.
[17] Hunter DJ. Gene-environment interactions in human diseases. Nat Rev Genet 2005;6(4):287–98.
[18] Zhou A, Liu X, Xia T, Li F, Wang J, Li J. Estrogen receptor alpha gene (ESR1) polymorphism and its interaction with smoking and drinking contribute to susceptibility of systemic lupus erythematosus. Immunol Res 2017;65(4):951–6.
[19] Kiyohara C, Washio M, Horiuchi T, Asami T, Ide S, Atsumi T, et al. Risk modification by CYP1A1 and GSTM1 polymorphisms in the association of cigarette smoking and systemic lupus erythematosus in a Japanese population. Scand J Rheumatol 2012;41(2):103–9.
[20] Kiyohara C, Washio M, Horiuchi T, Tada Y, Asami T, Ide S, et al. Cigarette smoking, N-acetyltransferase 2 polymorphisms and systemic lupus erythematosus in a Japanese population. Lupus 2009;18(7):630–8.
[21] Kiyohara C, Washio M, Horiuchi T, Tada Y, Asami T, Ide S, et al. Cigarette smoking, STAT4 and TNFRSF1B polymorphisms, and systemic lupus erythematosus in a Japanese population. J Rheumatol 2009;36(10):2195–203.
[22] Fraser PA, Ding WZ, Mohseni M, Treadwell EL, Dooley MA, St Clair EW, et al. Glutathione S-transferase M null homozygosity and risk of systemic lupus erythematosus associated with sun exposure: a possible gene-environment interaction for autoimmunity. J Rheumatol 2003;30(2):276–82.
[23] Kiyohara C, Washio M, Horiuchi T, Asami T, Ide S, Atsumi T, et al. Modifying effect of N-acetyltransferase 2 genotype on the association between systemic lupus erythematosus and consumption of alcohol and caffeine-rich beverages. Arthritis Care Res (Hoboken) 2014;66(7):1048–56.
[24] Petri M, Orbai AM, Alarcon GS, Gordon C, Merrill JT, Fortin PR, et al. Derivation and validation of the systemic lupus international collaborating clinics classification criteria for systemic lupus erythematosus. Arthritis Rheum 2012;64(8):2677–86.
[25] Jacobsen S, Starklint H, Petersen J, Ullman S, Junker P, Voss A, et al. Prognostic value of renal biopsy and clinical variables in patients with lupus nephritis and normal serum creatinine. Scand J Rheumatol 1999;28(5):288–99.
[26] Duarte-Garcia A, Barr E, Magder LS, Petri M. Predictors of incident proteinuria among patients with SLE. Lupus Sci Med 2017;4(1):e000200.
[27] Tanha N, Hansen RB, Nielsen CT, Faurschou M, Jacobsen S. Clinical and serological associations with the development of incident proteinuria in Danish patients with systemic lupus erythematosus. J Rheumatol 2018;45(7):934–41.

[28] Schreiber K, Sciascia S, de Groot PG, Devreese K, Jacobsen S, Ruiz-Irastorza G, et al. Antiphospholipid syndrome. Nat Rev Dis Primers 2018;4:18005.
[29] Arbuckle MR, McClain MT, Rubertone MV, Scofield RH, Dennis GJ, James JA, et al. Development of autoantibodies before the clinical onset of systemic lupus erythematosus. N Engl J Med 2003;349(16):1526–33.
[30] Eriksson C, Rantapaa-Dahlqvist S. Cytokines in relation to autoantibodies before onset of symptoms for systemic lupus erythematosus. Lupus 2014;23(7):691–6.
[31] Md Yusof MY, Psarras A, El-Sherbiny YM, Hensor EMA, Dutton K, Ul-Hassan S, et al. Prediction of autoimmune connective tissue disease in an at-risk cohort: prognostic value of a novel two-score system for interferon status. Ann Rheum Dis 2018;77(10):1432–9.
[32] Kahlenberg JM, Kaplan MJ. The inflammasome and lupus: another innate immune mechanism contributing to disease pathogenesis? Curr Opin Rheumatol 2014;26(5):475–81.
[33] Kawai T, Akira S. The role of pattern-recognition receptors in innate immunity: update on toll-like receptors. Nat Immunol 2010;11(5):373–84.
[34] Rasmussen NS, Jacobsen S. Microparticles - culprits in the pathogenesis of systemic lupus erythematosus? Expert Rev Clin Immunol 2018;14(6):443–5.
[35] Boudreau LH, Duchez AC, Cloutier N, Soulet D, Martin N, Bollinger J, et al. Platelets release mitochondria serving as substrate for bactericidal group IIA-secreted phospholipase A2 to promote inflammation. Blood 2014;124(14):2173–83.
[36] Ohlenschlaeger T, Garred P, Madsen HO, Jacobsen S. Mannose-binding lectin variant alleles and the risk of arterial thrombosis in systemic lupus erythematosus. N Engl J Med 2004;351(3):260–7.
[37] Hermansen ML, Lindhardsen J, Torp-Pedersen C, Faurschou M, Jacobsen S. The risk of cardiovascular morbidity and cardiovascular mortality in systemic lupus erythematosus and lupus nephritis: a Danish nationwide population-based cohort study. Rheumatology (Oxford, England) 2017;56(5):709–15.
[38] Faurschou M, Starklint H, Halberg P, Jacobsen S. Prognostic factors in lupus nephritis: diagnostic and therapeutic delay increases the risk of terminal renal failure. J Rheumatol 2006;33(8):1563–9.
[39] Leng RX, Wang W, Cen H, Zhou M, Feng CC, Zhu Y, et al. Gene-gene and gene-sex epistatic interactions of MiR146a, IRF5, IKZF1, ETS1 and IL21 in systemic lupus erythematosus. PLoS One 2012;7(12):e51090.
[40] Ceccarelli F, Perricone C, Borgiani P, Ciccacci C, Rufini S, Cipriano E, et al. Genetic factors in systemic lupus erythematosus: contribution to disease phenotype. J Immunol Res 2015;2015:745647.
[41] Jonsson S, Sveinbjornsson G, de Lapuente Portilla AL, Swaminathan B, Plomp R, Dekkers G, et al. Identification of sequence variants influencing immunoglobulin levels. Nat Genet 2017;49(8):1182–91.
[42] Lee-Kirsch MA, Gong M, Chowdhury D, Senenko L, Engel K, Lee YA, et al. Mutations in the gene encoding the 3′-5' DNA exonuclease TREX1 are associated with systemic lupus erythematosus. Nat Genet 2007;39(9):1065–7.
[43] Walport MJ, Davies KA, Botto M. C1q and systemic lupus erythematosus. Immunobiology 1998;199(2):265–85.
[44] Teruel M, Chamberlain C, Alarcon-Riquelme ME. Omics studies: their use in diagnosis and reclassification of SLE and other systemic autoimmune diseases. Rheumatology (Oxford) 2017;56(suppl_1):i78–87.
[45] Deng Y, Zhao J, Sakurai D, Kaufman KM, Edberg JC, Kimberly RP, et al. MicroRNA-3148 modulates allelic expression of toll-like receptor 7 variant associated with systemic lupus erythematosus. PLoS Genet 2013;9(2):e1003336.
[46] Enevold C, Nielsen CH, Jacobsen RS, Hermansen ML, Molbo D, Avlund K, et al. Single nucleotide polymorphisms in genes encoding toll-like receptors 7, 8 and 9 in Danish patients with systemic lupus erythematosus. Mol Biol Rep 2014;41(9):5755–63.
[47] Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW, et al. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. N Engl J Med 2007;357(10):977–86.
[48] Svenungsson E, Gustafsson J, Leonard D, Sandling J, Gunnarsson I, Nordmark G, et al. A STAT4 risk allele is associated with ischaemic cerebrovascular events and anti-phospholipid antibodies in systemic lupus erythematosus. Ann Rheum Dis 2010;69(5):834–40.
[49] Hagberg N, Joelsson M, Leonard D, Reid S, Eloranta ML, Mo J, et al. The STAT4 SLE risk allele rs7574865[T] is associated with increased IL-12-induced IFN-gamma production in T cells from patients with SLE. Ann Rheum Dis 2018;77(7):1070–7.
[50] Namjou B, Kim-Howard X, Sun C, Adler A, Chung SA, Kaufman KM, et al. PTPN22 association in systemic lupus erythematosus (SLE) with respect to individual ancestry and clinical sub-phenotypes. PLoS One 2013;8(8):e69404.
[51] Morris DL, Taylor KE, Fernando MM, Nititham J, Alarcon-Riquelme ME, Barcellos LF, et al. Unraveling multiple MHC gene associations with systemic lupus erythematosus: model choice indicates a role for HLA alleles and non-HLA genes in Europeans. Am J Hum Genet 2012;91(5):778–93.
[52] Fernando MM, Stevens CR, Sabeti PC, Walsh EC, McWhinnie AJ, Shah A, et al. Identification of two independent risk factors for lupus within the MHC in United Kingdom families. PLoS Genet 2007;3(11):e192.
[53] Graham RR, Ortmann WA, Langefeld CD, Jawaheer D, Selby SA, Rodine PR, et al. Visualizing human leukocyte antigen class II risk haplotypes in human systemic lupus erythematosus. Am J Hum Genet 2002;71(3):543–53.
[54] Wild CP. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol Biomarkers Prev 2005;14(8):1847–50.
[55] Rappaport SM, Smith MT. Epidemiology. Environment and disease risks. Science 2010;330(6003):460–1.
[56] Costenbader KH, Kim DJ, Peerzada J, Lockman S, Nobles-Knight D, Petri M, et al. Cigarette smoking and the risk of systemic lupus erythematosus: a meta-analysis. Arthritis Rheum 2004;50(3):849–57.
[57] Finckh A, Cooper GS, Chibnik LB, Costenbader KH, Watts J, Pankey H, et al. Occupational silica and solvent exposures and risk of systemic lupus erythematosus in urban women. Arthritis Rheum 2006;54(11):3648–54.
[58] Costenbader KH, Feskanich D, Stampfer MJ, Karlson EW. Reproductive and menopausal factors and risk of systemic lupus erythematosus in women. Arthritis Rheum 2007;56(4):1251–62.
[59] Wang J, Pan HF, Ye DQ, Su H, Li XP. Moderate alcohol drinking might be protective for systemic lupus erythematosus: a systematic review and meta-analysis. Clin Rheumatol 2008;27(12):1557–63.
[60] Gomes JP, Watad A, Shoenfeld Y. Nicotine and autoimmunity: the lotus' flower in tobacco. Pharmacol Res 2018;128:101–9.
[61] Parks CG, De Roos AJ. Pesticides, chemical and industrial exposures in relation to systemic lupus erythematosus. Lupus 2014;23(6):527–36.
[62] Shen S, Zhang R, Zhang J, Wei Y, Guo Y, Su L, et al. Welding fume exposure is associated with inflammation: a global metabolomics profiling study. Environ Health 2018;17(1):68.
[63] Ulff-Moller CJ, Nielsen NM, Rostgaard K, Hjalgrim H, Frisch M. Epstein-Barr virus-associated infectious mononucleosis and risk of systemic lupus erythematosus. Rheumatology (Oxford) 2010;49(9):1706–12.
[64] Rasmussen NS, Draborg AH, Nielsen CT, Jacobsen S, Houen G. Antibodies to early EBV, CMV, and HHV6 antigens in systemic lupus erythematosus patients. Scand J Rheumatol 2015;44(2):143–9.
[65] Draborg AH, Duus K, Houen G. Epstein-Barr virus and systemic lupus erythematosus. Clin Dev Immunol 2012;2012:370516.
[66] Barbhaiya M, Costenbader KH. Ultraviolet radiation and systemic lupus erythematosus. Lupus 2014;23(6):588–95.
[67] Hedrich CM. Epigenetics in SLE. Curr Rheumatol Rep 2017;19(9):58.
[68] Jeffries MA, Dozmorov M, Tang Y, Merrill JT, Wren JD, Sawalha AH. Genome-wide DNA methylation patterns in CD4+ T cells from patients with systemic lupus erythematosus. Epigenetics 2011;6(5):593–601.
[69] Absher DM, Li X, Waite LL, Gibson A, Roberts K, Edberg J, et al. Genome-wide DNA methylation analysis of systemic lupus erythematosus reveals persistent hypomethylation of interferon genes and compositional changes to CD4+ T-cell populations. PLoS Genet 2013;9(8):e1003678.
[70] Coit P, Jeffries M, Altorok N, Dozmorov MG, Koelsch KA, Wren JD, et al. Genome-wide DNA methylation study suggests epigenetic accessibility and transcriptional poising of interferon-regulated genes in naive CD4+ T cells from lupus patients. J Autoimmun 2013;43:78–84.
[71] Coit P, Yalavarthi S, Ognenovski M, Zhao W, Hasni S, Wren JD, et al. Epigenome profiling reveals significant DNA demethylation of interferon signature genes in lupus neutrophils. J Autoimmun 2015;58:59–66.
[72] Ulff-Moller CJ, Asmar F, Liu Y, Svendsen AJ, Busato F, Gronbaek K, et al. Twin DNA methylation profiling reveals flare-dependent interferon signature and B cell promoter hypermethylation in systemic lupus erythematosus. Arthritis Rheumatol 2018;70(6):878–90.
[73] Javierre BM, Fernandez AF, Richter J, Al-Shahrour F, Martin-Subero JI, Rodriguez-Ubreva J, et al. Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. Genome Res 2010;20(2):170–9.
[74] Yeung KS, Chung BH, Choufani S, Mok MY, Wong WL, Mak CC, et al. Genome-wide DNA methylation analysis of Chinese patients with systemic lupus erythematosus identified hypomethylation in genes related to the type I interferon pathway. PLoS One 2017;12(1):e0169553.
[75] Chung SA, Nititham J, Elboudwarej E, Quach HL, Taylor KE, Barcellos LF, et al. Genome-wide assessment of differential DNA methylation associated with autoantibody production in systemic lupus erythematosus. PLoS One 2015;10(7):e0129813.
[76] Liu J, Morgan M, Hutchison K, Calhoun VD. A study of the influence of sex on genome wide methylation. PLoS One 2010;5(4):e10028.
[77] Heyn H, Moran S, Hernando-Herraez I, Sayols S, Gomez A, Sandoval J, et al. DNA methylation contributes to natural human variation. Genome Res 2013;23(9):1363–72.
[78] Ambatipudi S, Cuenin C, Hernandez-Vargas H, Ghantous A, Le Calvez-Kelm F, Kaaks R, et al. Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study. Epigenomics 2016;8(5):599–618.
[79] Abdul QA, Yu BP, Chung HY, Jung HA, Choi JS. Epigenetic modifications of gene expression by lifestyle and environment. Arch Pharm Res 2017;40(11):1219–37.
[80] Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, et al. Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci U S A 2005;102(30):10604–9.
[81] Zhao M, Liu S, Luo S, Wu H, Tang M, Cheng W, et al. DNA methylation and mRNA and microRNA expression of SLE CD4+ T cells correlate with disease phenotype. J Autoimmun 2014;54:127–36.
[82] Zhou Y, Qiu X, Luo Y, Yuan J, Li Y, Zhong Q, et al. Histone modifications and methyl-CpG-binding domain protein levels at the TNFSF7 (CD70) promoter in SLE CD4+ T cells. Lupus 2011;20(13):1365–71.
[83] Harley JB, Chen X, Pujato M, Miller D, Maddox A, Forney C, et al. Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. Nat Genet 2018;50(5):699–707.
[84] Ottman R. Gene-environment interaction: definitions and study designs. Prev Med 1996;25(6):764–70.
[85] Ottman R. An epidemiologic approach to gene-environment interaction. Genet Epidemiol 1990;7(3):177–85.
[86] Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature 2008;452(7187):638–42.

[87] Strickland FM, Hewagama A, Wu A, Sawalha AH, Delaney C, Hoeltzel MF, et al. Diet influences expression of autoimmune-associated genes and disease severity by epigenetic mechanisms in a transgenic mouse model of lupus. Arthritis Rheum 2013;65(7):1872–81.
[88] Belfield KD, Tichy EM. Review and drug therapy implications of glucose-6-phosphate dehydrogenase deficiency. Am J Health Syst Pharm 2018;75(3):97–104.
[89] Rothman KJ. Epidemiology. An introduction. 2nd ed. New York: Oxford University Press, Inc.; 2012.
[90] Ahlbom A, Alfredsson L. Interaction: a word with two meanings creates confusion. Eur J Epidemiol 2005;20(7):563–4.
[91] Rothman KJ. Causes. Am J Epidemiol 1976;104(6):587–92.
[92] Andersson T, Alfredsson L, Kallberg H, Zdravkovic S, Ahlbom A. Calculating measures of biological interaction. Eur J Epidemiol 2005;20(7):575–9.
[93] Sparks JA, Costenbader KH. Genetics, environment, and gene-environment interactions in the development of systemic rheumatic diseases. Rheum Dis Clin North Am 2014;40(4):637–57.
[94] Thomas D. Gene–environment-wide association studies: emerging approaches. Nat Rev Genet 2010;11(4):259–72.
[95] Perry Jr. HM, Tan EM, Carmody S, Sakamoto A. Relationship of acetyl transferase activity to antinuclear antibodies and toxic symptoms in hypertensive patients treated with hydralazine. J Lab Clin Med 1970;76(1):114–25.
[96] Woosley RL, Drayer DE, Reidenberg MM, Nies AS, Carr K, Oates JA. Effect of acetylator phenotype on the rate at which procainamide induces antinuclear antibodies and the lupus syndrome. N Engl J Med 1978;298(21):1157–9.
[97] Nilsson S, Makela S, Treuter E, Tujague M, Thomsen J, Andersson G, et al. Mechanisms of estrogen action. Physiol Rev 2001;81(4):1535–65.
[98] Locksley RM, Killeen N, Lenardo MJ. The TNF and TNF receptor superfamilies: integrating mammalian biology. Cell 2001;104(4):487–501.
[99] Parascandola M, Weed DL. Causation in epidemiology. J Epidemiol Community Health 2001;55(12):905–12.
[100] Rothman KJ, Greenland S. Causation and causal inference in epidemiology. Am J Public Health 2005;95(Suppl. 1):S144–50.
[101] Hill AB. The environment and disease: association or causation? 1965. J R Soc Med 2015;108(1):32–7.
[102] Klareskog L, Catrina AI, Paget S. Rheumatoid arthritis. Lancet 2009;373(9664):659–72.
[103] Klareskog L, Stolt P, Lundberg K, Kallberg H, Bengtsson C, Grunewald J, et al. A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. Arthritis Rheum 2006;54(1):38–46.
[104] Pedersen M, Jacobsen S, Garred P, Madsen HO, Klarlund M, Svejgaard A, et al. Strong combined gene-environment effects in anti-cyclic citrullinated peptide-positive rheumatoid arthritis: a nationwide case-control study in Denmark. Arthritis Rheum 2007;56(5):1446–53.
[105] Kong A, Masson G, Frigge ML, Gylfason A, Zusmanovich P, Thorleifsson G, et al. Detection of sharing by descent, long-range phasing and haplotype imputation. Nat Genet 2008;40(9):1068–75.
[106] Coombes B, Basu S, McGue M. A combination test for detection of gene-environment interaction in cohort studies. Genet Epidemiol 2017;41(5):396–412.
[107] Lin X, Lee S, Christiani DC, Lin X. Test for interactions between a genetic marker set and environment in generalized linear models. Biostatistics 2013;14(4):667–81.
[108] McCue ME, McCoy AM. The scope of big data in one medicine: unprecedented opportunities and challenges. Front Vet Sci 2017;4:194.
[109] Murcray CE, Lewinger JP, Conti DV, Thomas DC, Gauderman WJ. Sample size requirements to detect gene-environment interactions in genome-wide association studies. Genet Epidemiol 2011;35(3):201–10.
[110] Sturmer T, Brenner H. Flexible matching strategies to increase power and efficiency to detect and estimate gene-environment interactions in case-control studies. Am J Epidemiol 2002;155(7):593–602.
[111] Arnold BF, Hogan DR, Colford Jr. JM, Hubbard AE. Simulation methods to estimate design power: an overview for applied research. BMC Med Res Methodol 2011;11:94.
[112] Jeong DY, Lee SW, Park YH, Choi JH, Kwon YW, Moon G, et al. Genetic variation and systemic lupus erythematosus: a field synopsis and systematic meta-analysis. Autoimmun Rev 2018;17(6):553–66.
[113] Pedersen M, Jacobsen S, Klarlund M, Frisch M. Socioeconomic status and risk of rheumatoid arthritis: a Danish case-control study. J Rheumatol 2006;33(6):1069–74.